(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 524 385 A2**

(12) **EUROPEAN PATENT APPLICATION**

(21) Application number: **92107831.7**

(51) Int. Cl.5: **G06F 15/20**

(22) Date of filing: **09.05.92**

(30) Priority: **05.06.91 US 710638**

(43) Date of publication of application:
27.01.93 Bulletin 93/04

(84) Designated Contracting States:
DE FR GB

(71) Applicant: **International Business Machines Corporation**
Old Orchard Road
Armonk, N.Y. 10504(US)

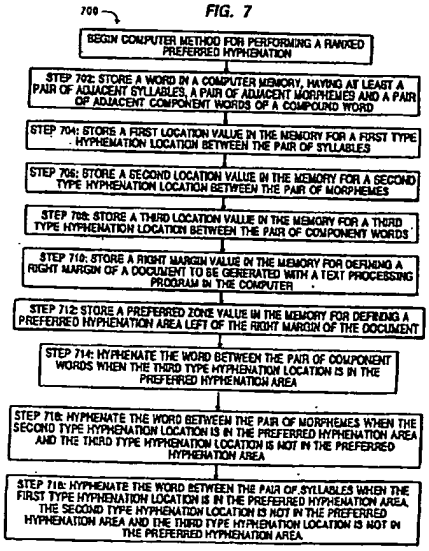
(72) Inventor: **Hersey, Ian L.**
6213 Wagner Lane
Bethesba, MD 20816(US)
Inventor: **Stephens, Richard L.**
7504 Filbert Terrace
Galthersburg, MD 20816(US)
Inventor: **Zamora, Antonio**
4601 North Park Avenue
No. 411 Chevy Chase, MD 20817(US)

(74) Representative: **Jost, Ottokarl, Dipl.-Ing.**
IBM Deutschland GmbH Patentwesen und Urheberrecht Schönaicher Strasse 220
W-7030 Böblingen(DE)

(54) **Computer method for ranked hyphenation of multilingual text.**

(57) The hierarchy of the hyphenation points in a word is considered and the amount of space remaining in the line where the hyphenated word is to fit is determined, thus selecting the best hyphenation point for the word. Instead of selecting the last possible hyphen within the hot zone (HZ), the hot zone (HZ) is scanned to find hyphenation points that may be better than the last hyphen within the hot zone (HZ). The higher a hyphenation point is in the hierarchy, the more preferred is the hyphen.

FIG. 7



EP 0 524 385 A2

Background of the Invention

1. Technical Field

The invention disclosed broadly relates to computer methods, and particularly relates to a computer method for choosing hyphenation points in multilingual text.

2. Background Art

Mechanized hyphenation is a process that is necessary for computerized word processing and printing applications. Mechanized hyphenation has been attempted by the use of stored dictionary systems, rule-based systems, and statistical systems. Dictionary-based systems store the hyphenation for each dictionary entry. Rule-based systems use rules that may apply to more than one word. Rules may be non-specific and apply to any word, or they may be associated with specific words to provide their hyphenation points. Finally, the statistical systems use tables of hyphenation statistics based on a collection of words and apply these statistics to determine the hyphenation of other words. Some of these statistical techniques insert numbers within a word to be hyphenated that indicate the confidence with which the word can be hyphenated at that particular point.

Carlgren describes a system that combines several techniques in IBM Technical Disclosure Bulletin, Vol. 26, No. 11, pp. 6108-6109 and pp. 6095-6096. Carlgren elaborates on the combination of dictionary-based hyphenation and rule-based hyphenation in USP 4,574,363. Rosenbaum describes dictionary-based methods of hyphenation in USPs 4,028,677 and 4,092,729. Herzik also describes a dictionary-based method for hyphenating multilingual text in USP 4,456,969. Zamora in USP 5,113,342 describes a rule-based method for hyphenating text.

Another prior art technique by Moore in the IBM Technical Disclosure Bulletin, Vol. 29, No. 1, pp. 383-384 combines manual selection of hyphenation points and dictionary hyphenation.

Objects of the Invention

An object of the invention is to provide an improved dictionary that makes it possible to rank the importance of each hyphenation point in a word.

Another object of the invention is to provide an improved computer process to select the linguistically best hyphenation point in a word.

A further object of the invention is to provide a mechanism for dehyphenating previously hyphenated words.

Summary of the Invention

These and other objects are accomplished by the disclosed invention. A computer method is disclosed for performing a ranked preferred hyphenation of words. The method includes the step of storing a word in a computer memory, having at least a pair of adjacent syllables, a pair of adjacent morphemes and a pair of adjacent component words. The method continues by storing a first location value in the memory for a first type hyphenation location between the pair of syllables. The method continues by storing a second location value in the memory for a second type hyphenation location between the pair of morphemes. The method continues by storing a third location value in the memory for a third type hyphenation location between the pair of component words. The method continues by storing a right margin value in the memory for defining a right margin of a document to be generated with a text processing program in the computer.

The method continues by storing a preferred zone value in the memory for defining a preferred hyphenation area left of the right margin of the document. The method continues by hyphenating the word between the pair of component words when the third type hyphenation location is in the preferred hyphenation area. The method continues by hyphenating the word between the pair of morphemes when the second type hyphenation location is in the preferred hyphenation area and the third type hyphenation location is not in the preferred hyphenation area.

Finally, the method continues by hyphenating the word between the pair of syllables when the first type hyphenation location is in the preferred hyphenation area, the second type hyphenation location is not in the preferred hyphenation area and the third type hyphenation location is not in the preferred hyphenation area. In this manner the best hyphenation point for the word is selected.

WEST

Description of the Figures

The invention will be more fully appreciated by referring to the accompanying figures.

Fig. 1 shows a Text Processing System with Ranked Preferred Hyphenation.
Fig. 2 shows a Hyphen Type Ranking.
Fig. 3 shows a Example of Three Types of Soft Hyphenation.
Fig. 4 shows a Comparison of Preferred Ranked and Traditional Hyphenation Methods.
Fig. 5 shows a Example of Hard Hyphen Character and Three Soft Hyphens.
Fig. 6 shows a Example of Preferred Hyphenation Method on Word with Hard Hyphen Character and Three Soft Hyphens.
Fig. 7 shows a Flow Diagram for Hyphenating a Word with Three Types of Ranked Soft Hyphens.
Fig. 8 shows a Flow Diagram for Hyphenating a word with Three Types of Ranked Soft Hyphens And a Hard Hyphen.
Fig. 9 shows a Special Soft Hyphen Processing for German Words.
Fig. 10 shows a Dehyphenation for Hyphenated Special Processing German Words.
Fig. 11 shows a Flow Diagram of Preferred Ranked Special Processing Hyphenation.

Discussion of the Preferred Embodiment

The invention consists of a manually-constructed dictionary that identifies hyphenation points and their types and a process for selecting the best hyphenation point under particular circumstances. The dictionary entries contain the traditional hard hyphens (-) and soft hyphens (_) so that a word like "high-spirited" is marked as "high-spir_it_ed." In addition to these marks the equal (=) indicates a preferred hyphen, a plus (+) indicates a component hyphen, a comma (,) indicates a hyphenation point requiring special processing. The hierarchy of these marks in decreasing preference is: 1) hard hyphen; 2) component hyphen; 3) special processing hyphen; 4) preferred hyphen; and 5) soft hyphen. The first three types generally delimit word component boundaries, the preferred hyphen usually marks morpheme boundaries, and the soft hyphen separates syllable boundaries.

The process of automatically hyphenating a word requires an input parameter called the "hot zone" and consisting of a number that indicates how much latitude the program has for selecting a hyphenation point. If the number is 5, for example, the program will try to find a hyphenation point within the five characters of the word that fall into the "hot zone." The usual procedure in the prior art techniques is to choose the last possible hyphen within the hot zone. However, this often produces inappropriate hyphenation.

This invention improves upon the prior art by taking into consideration the hierarchy of the hyphenation points in the word and the amount of space remaining in the line where the hyphenated word is to fit, thus selecting the best hyphenation point for the word. Instead of selecting the last possible hyphen within the hot zone, this invention scans the hot zone and a small portion preceding the hot zone to find hyphenation points that may be better than the last hyphen within the hot zone. The higher a hyphenation point is in the hierarchy described above, the better the hyphen is.

A complementary function to hyphenation is dehyphenation. Dehyphenation is the removal of hyphens. This process is necessary when hyphenated text is changed in a word processing system and previously hyphenated components of a word need to be put back together. In general, dehyphenation is a process symmetrical to hyphenation. The language-specific exceptions to dehyphenation and the specific rules for hyphenation points requiring special processing (those marked with a comma) are described below. In addition, the hyphenation process incorporates traditional hyphenation rules for the following languages:

CATALAN

A high dot that separates a double "l" disappears when the word is hyphenated at that point. Thus, "novel.la" becomes "novel-/la" where the slash represents a line break.

DUTCH

When a word is hyphenated preceding a vowel with a dieresis, the dieresis is removed. The word "geïsoleerd" when hyphenated after "ge" loses the dieresis and becomes "ge-/isoleerd." Words such as this are marked in the dictionary with a comma at that point: "ge,isoleerd" and it is necessary to translate the character with dieresis to the character without dieresis in the corresponding upper or lower case.

Another special case is the elimination of a double letter, e.g., "chocolaatje" hyphenates as "chocola-

WEST

/tje." The dictionary entry for this word is marked as "chocola,atje." Words with identical vowels on either side of the special hyphenation marker lose the second vowel upon hyphenation except when the vowel is "e." For the cases with "e,e," the following rule is used: "e,etje" -- "é-/tje," as in the following words:

cafeetje	café-tje
clicheetje	cliché-tje
comiteetje	comité-tje

There are three exceptions to this rule (six if you count the plural forms) which must be handled differently. These are:

dejeuneetje	dejeuner-tje
dineetje	diner-tje
soupeetje	souper-tje

Dehyphenation of these specially-processed words requires application of the following transformation rules:

Hyphen context	Transformation	Example
a-a	aä	naäpen
a-e	aë	pentaëder
a-i	aï	ma'ïzena
a-u	aü	Kafarnaüm
e-e	eë	geërfd
e-ij	eij	beijverd
e-i	eï	beïnvloeden
e-ui	eui	geuit
e-u	eü	geürineerd
i-ee	iee	officieel
i-eu	ieu	officieus
i-e	ië	officiële
o-e	oë	zoëven
o-i	oï	colloïdaal
o-o	oö	zoölogie
u-i	uï	fluïdum
u-u	uü	vacuüm
a-tje	aatje	chocolaatje
é-tje	eetje	employeetje
o-tje	ootje	autootje
u-tje	uutje	menuutje
dejeuner-tje	dejeuneetje	dejeuneetjes
diner-tje	dineetje	dineetjes
souper-tje	soupeetje	soupeetjes

GERMAN

A soft hyphen between "c" and "k" requires changing the "c" to another "k." The Germany word "Zuc_ker" (sugar) is hyphenated "Zuk-/ker," where the "ck" changes to "k-k."

Germany requires special processing for compound words with letters that have been elided. In general, these words involve a double consonant and a vowel, e.g., "Schiffahrt" ("navigation") is hyphenated as "Schiff-/fahrt." The dictionary entry for this word is marked as "Schif,fahrt."

PORTUGUESE

When a word is hyphenated at a hard hyphen another hyphen is added at the beginning of the next line. Thus, the Portuguese word "guarda-chuva" ("umbrella") is hyphenated as a "guarda-/-chuva."

The dictionary for specifying hyphenation can be extended by the use of hyphenation rules for endings. This can save substantial storage space. Consider, for example, the word "arm=bånds+ur" (Danish for "watch") which has a preferred hyphen point after "arm" (shown here by an equal sign), and a component hyphen point before the final "ur" (shown as a plus sign).

By giving only the hyphenation positions and types for one form of the word, the hyphenation for variants, such as plurals, can be derived from application of rules associated with each suffix. Since the suffixes may apply to many types of words, it is necessary to give priority to the hyphenation types indicated in the dictionary when hyphenating at the corresponding positions. For example, the word "armbåndsure" which adds an "e" ending to the stem "armbåndsur," is associated with a rule "__u__," which would give a hyphenation "arm=bånds__u__re." However, because the position between the "s" and the "u" is "+" (compound) in the dictionary, the resulting hyphenation must be "arm=bånds+u__re."

Fig. 1 shows Text Processing System with Ranked Preferred Hyphenation in accordance with the invention. CPU 20 is connected by Bus 22 to the Memory 24, the Display and keyboard 26, the printer 28, the communications adapter 30 and the DASD 31.

Memory 24 is partitioned into the Dictionary 32, the Ranked Preferred Hyphenation Program 34, the Special Processing Soft Hyphen Subroutines 36, the Text Processing Program 38 and the Operating System 40. The CPU 20 executes the programmed executable statements of the programs 34, 36, 38, and 40 .In the memory 24, operating on text input from the keyboard 26 and displayed on the display, the finished text then being output to the printer 28, the communications adapter 30 or the DASD 31.

The ranked preferred hyphenation program 34 is a programmed sequence of executable statements which embody the computer methods depicted in the flow diagrams of Figs. 7, 8 and 11. The method 700 of Fig. 7, is a computer method for Hyphenating a Word with Three Types of Ranked Soft Hyphens. The method 800 of Fig. 8 is a computer method for Hyphenating a word with Three Types of Ranked Soft Hyphens And a Hard Hyphen. The method 1100 of Fig. 11 is a computer method for Preferred Ranked Special Processing Hyphenation.

The special processing soft hyphen subroutines 36 are programmed sequences of executable statements which embody computer methods for hyphenating the word between the pair of special processing segments and performing special processing to at least one segment of the pair of segments when the second type hyphenation location is in the preferred hyphenation area. The special processing steps can be any of those previously discussed herein, for German, Catalan, Dutch, Portuguese and other languages. An example is given in Fig. 9 which shows the Special Soft Hyphen Processing for German Words. The reverse process of dehyphenation is also embodied in the programmed sequences of the subroutines 36. For example, Fig. 10 shows the Dehyphenation for Hyphenated Special Processing German Words, which can occur when reflowing a paragraph.

The Dictionary 32 includes a plurality of words in the Dictionary Word Field, each word associated with its own Hyphenation Mask field and Hyphenation Type Field.

In Fig. 1, Dictionary Word Field includes words such as "Photomicroscopic" and "Superabundant". Associated with the word "Photomicroscopic" is its hyphenation mask "3 2 2 3 4". The first number "3" represents an offset from the beginning of the word to the first hyphenation point between "o" and "t". Each number thereafter represents the offset from the preceding hyphenation point to the next hyphenation point.

Also associated with the word "Photomicroscopic" is its hyphenation type field "__ + __ = __". Each symbol represents the rank of the corresponding hyphen located by the hyphenation mask field for the word.

Similarly the word "Superabundant" has a Hyphenation Mask field and a Hyphenation Type Field of "2 3 1 3" and "__ + __ __", respectively.

Fig. 2 shows the Hyphen Type Rank. The Hyphen Type Rank from High rank to Low rank is as follows.

- is a hard hyphen (ASCII character 045), the highest rank.
+ is a component word boundary soft hyphen for a compound word.
, is a special processing soft hyphen (non-English).
= is a morpheme boundary soft hyphen.
__ is a syllable boundary soft hyphen, the lowest rank.

Fig. 3 shows an Example of Three Types of Soft Hyphenation for the word "photomicroscopic". "Pho__to+mi__cro=scop__ic" has as its Hyphenation Mask " 3 2 2 3 4 " and as its Hyphen Type Mask "__ + __ = __".

Fig. 4 shows a Comparison of Preferred Ranked and Traditional Hyphenation Methods. The hot zone is

designated HZ, the preferred processing zone is designated PZ and the right margin is designated RM.

The hot zone HZ has the function defining the left boundary of the region between itself and the right margin PM where traditional hyphenation processing takes place. If the function of preferred ranked hyphenation is disabled for a particular application, the hot zone HZ will define the area where all soft hyphenation will take place, the process being performed on the nearest soft hyphenation point to the right margin RM to split the word in the traditional hyphenation process. However, if the function of preferred ranked hyphenation is enabled for a particular application, in accordance with the invention, then the preferred zone PZ will define the left boundary of the area between itself and the right margin RM where the ranked hyphenation process will take place, the inventive process being performed on the highest ranking hyphenation point within the preferred zone PZ. The preferred zone PZ can be defined by the user to be located at any column in the document working area, and it can be coincident with the hot zone HZ as is shown in Fig. 6 or PZ can be separate from HZ as is shown in Fig. 4.

Note the difference between the preferred ranked method of the invention and the traditional method for the words on line b and b' in Fig. 4. "Photomicro" is the hyphenation resulting from the invention, which divides the word at the morpheme boundary "micro". In the traditional method of hyphenation, the last available hyphenation point is selected, which is the syllable boundary "scop", which is not a linguistically preferred location for dividing the word.

Note again the lines g, h, and i in Fig. 4, for the preferred ranked method of the invention and compare this with the traditional method of lines g', h', and i', respectively. The preferred ranked method of the invention divides the word at the compound word component word boundary "photo". In the traditional method of hyphenation, the last available hyphenation point is selected, which is the syllable boundary "photomi", which is not a linguistically preferred location for dividing the word.

Fig. 5 shows an Example of Hard Hyphen Character and Three Soft Hyphens. The word "photomicroscopic-oxbow" is a coined word for the oxbow shape of an object depicted in a photomicrograph. Its ranked hyphenation points are "Pho__to + mi__cro = scop__ic-ox + bow".

Fig. 6 is an Example of Preferred Hyphenation Method on a Word with a Hard Hyphen Character and Three Soft Hyphens, for example "photomicroscopic-oxbow". Note that in line b, the word is hyphenated at the preferred hard hyphen "photomicroscopic", in accordance with the invention, instead of at the last available syllable "photomicroscopic-ox", which would be the traditional method of hyphenation.

Fig. 7 is a Flow Diagram for Hyphenating a Word with Three Types of Ranked Soft Hyphens.

The method 700 is A computer method for performing a ranked preferred hyphenation of words.

The method performs step 702 storing a word in a computer memory, having at least a pair of adjacent syllables, a pair of adjacent morphemes and a pair of adjacent component words of a compound word.

The method then performs step 704 storing a first location value in the memory for a first type hyphenation location between the pair of syllables.

The method then performs step 706 storing a second location value in the memory for a second type hyphenation location between the pair of morphemes.

The method then performs step 708 storing a third location value in the memory for a third type hyphenation location between the pair of component words.

The method then performs step 710 storing a right margin value in the memory for defining a right margin of a document to be generated with a text processing program in the computer.

The method then performs step 712 storing a preferred zone value in the memory for defining a preferred hyphenation area left of the right margin of the document.

The method then performs step 714 hyphenating the word between the pair of component words when the third type hyphenation location is in the preferred hyphenation area.

The method then performs step 716 hyphenating the word between the pair of morphemes when the second type hyphenation location is in the preferred hyphenation area and the third type hyphenation location is not in the preferred hyphenation area.

The method then performs step 718 hyphenating the word between the pair of syllables when the first type hyphenation location is in the preferred hyphenation area, the second type hyphenation location is not in the preferred hyphenation area and the third type hyphenation location is not in the preferred hyphenation area.

Fig. 8 is a Flow Diagram for Hyphenating a word with Three Types of Ranked Soft Hyphens And a Hard Hyphen.

The method 800 is a computer method for performing a ranked preferred hyphenation of words.

The method performs step 802 storing a word in a computer memory, having at least a pair of adjacent syllables, a pair of adjacent morphemes, a pair of adjacent component words of a compound word and a hard hyphen.

WEST

The method then performs step 804 storing a first location value in the memory for a first type hyphenation location between the pair of syllables.

The method then performs step 806 storing a second location value in the memory for a second type hyphenation location between the pair of morphemes.

The method then performs step 808 storing a third location value in the memory for a third type hyphenation location between the pair of component words of a compound word.

The method then performs step 810 storing a right margin value in the memory for defining a right margin of a document to be generated with a text processing program in the computer.

The method then performs step 812 storing a preferred zone value in the memory for defining a preferred hyphenation area left of the right margin of the document.

The method then performs step 814 hyphenating the word at the hard hyphen when the hard hyphen is detected in the preferred hyphenation area.

The method then performs step 816 hyphenating the word between the pair of component words of a compound word when the third type hyphenation location is in the preferred hyphenation area and the hard hyphen is not detected in the preferred hyphenation area.

The method then performs step 818 hyphenating the word between the pair of morphemes when the second type hyphenation location is in the preferred hyphenation area, the third type hyphenation location is not in the preferred hyphenation area and the hard hyphen is not detected in the preferred hyphenation area.

The method then performs step 820 hyphenating the word between the pair of syllables when the first type hyphenation location is in the preferred hyphenation area, the second type hyphenation location is not in the preferred hyphenation area, the third type hyphenation location is not in the preferred hyphenation area and the hard hyphen is not detected in the preferred hyphenation area.

Fig. 9 shows the Special Soft Hyphen Processing for German Words.

Step 902 is: If "f,f" then "ff/f". The Dictionary in memory 24 has the entry "Schif,fahrt", and the preferred hyphenation adds an "f" to "Schiff" and places the remaining word "fahrt" on the next line.

Step 904 is: If "c,k" then "k/k". The dictionary in memory 24 has the entry "Zuc,ker", and the preferred hyphenation changes the "c" to a "k" for "Zuk" and places the remaining "ker" on the next line.

Fig. 10 shows the Dehyphenation for Hyphenated Special Processing German Words. This can occur when reflowing a paragraph.

Step 1002 is:

If "ff/f" then check the dictionary in memory 24. If the word "Schiff" is at the end of the first line and the word "fahrt" is at the beginning of the second line, then remove the "f" in "Schiff" and form the recombined word "Schiffahrt".

Step 1004 is:

If "k/k" then check the dictionary in memory 24. If the word "Zuk" is at the end of the first line and the word "ker" is at the beginning of the second line, then change the "k" to "c" in "Zuk" and form the recombined word "Zucker".

Fig. 11 is a Flow Diagram of Preferred Ranked Special Processing Hyphenation.

The method 1100 of Fig. 11 performs step 1101 storing a word in a computer memory, having at least a pair of adjacent syllables and a pair of adjacent special processing segments.

The method then performs step 1102 storing a first location value in the memory for a first type hyphenation location between the pair of syllables.

The method then performs step 1104 storing a second location value in the memory for a second type hyphenation location between the pair of special processing segments.

The method then performs step 1106 storing a right margin value in the memory for defining a right margin of a document to be generated with a text processing program in the computer.

The method then performs step 1108 storing a preferred zone value in the memory for defining a preferred hyphenation area left of the right margin of the document.

The method then performs step 1110 hyphenating the word between the pair of special processing segments and performing special processing to at least one segment of the pair of segments when the second type hyphenation location is in the preferred hyphenation area. The special processing steps can be any of those previously discussed herein, for German, Catalan, Dutch, Portuguese and other languages.

The method then performs step 1112 hyphenating the word between the pair of syllables when the first type hyphenation location is in the preferred hyphenation area and the second type hyphenation location is not in the preferred hyphenation area.

The resulting hyphenation is the most preferred for conveying the author's meaning to the reader.

WEST

Claims

1. A computer method for performing a ranked preferred hyphenation of words, comprising the steps of:

storing a word in a computer memory, having at least a pair of adjacent syllables and a pair of adjacent morphemes;

storing a first location value in said memory for a first type hyphenation location between said pair of syllables;

storing a second location value in said memory for a second type hyphenation location between said pair of morphemes;

storing a right margin value in said memory for defining a right margin of a document to be generated with a text processing program in said computer;

storing a preferred zone value in said memory for defining a preferred hyphenation area left of said right margin of said document;

hyphenating said word between said pair of morphemes when said second type hyphenation location is in said preferred hyphenation area;

hyphenating said word between said pair of syllables when said first type hyphenation location is in said preferred hyphenation area and said second type hyphenation location is not in said preferred hyphenation area.

2. A computer method for performing a ranked preferred hyphenation of words, comprising the steps of:

storing a word in a computer memory, having at least a pair of adjacent syllables and a pair of adjacent component words;

storing a first location value in said memory for a first type hyphenation location between said pair of syllables;

storing a second location value in said memory for a second type hyphenation location between said pair of component words;

storing a right margin value in said memory for defining a right margin of a document to be generated with a text processing program in said computer;

storing a preferred zone value in said memory for defining a preferred hyphenation area left of said right margin of said document;

hyphenating said word between said pair of component words when said second type hyphenation location is in said preferred hyphenation area;

hyphenating said word between said pair of syllables when said first type hyphenation location is in said preferred hyphenation area and said second type hyphenation location is not in said preferred hyphenation area.

3. A computer method for performing a ranked preferred hyphenation of words, comprising the steps of:

storing a word in a computer memory, having at least a pair of adjacent morphemes and a pair of adjacent component words;

storing a first location value in said memory for a first type hyphenation location between said pair of morphemes;

WEST

storing a second location value in said memory for a second type hyphenation location between said pair of component words;

storing a right margin value in said memory for defining a right margin of a document to be generated with a text processing program in said computer;

storing a preferred zone value in said memory for defining a preferred hyphenation area left of said right margin of said document;

hyphenating said word between said pair of component words when said second type hyphenation location is in said preferred hyphenation area;

hyphenating said word between said pair of morphemes when said first type hyphenation location is in said preferred hyphenation area and said second type hyphenation location is not in said preferred hyphenation area.

4. A computer method for performing a ranked preferred hyphenation of words, comprising the steps of:

storing (702) a word in a computer memory, having at least a pair of adjacent syllables, a pair of adjacent morphemes and a pair of adjacent component words;

storing (704) a first location value in said memory for a first type hyphenation location between said pair of syllables;

storing (706) a second location value in said memory for a second type hyphenation location between said pair of morphemes;

storing (708) a third location value in said memory for a third type hyphenation location between said pair of component words;

storing (710) a right margin value in said memory for defining a right margin of a document to be generated with a text processing program in said computer;

storing (712) a preferred zone value in said memory for defining a preferred hyphenation area left of said right margin of said document;

hyphenating (714) said word between said pair of component words when said third type hyphenation location is in said preferred hyphenation area;

hyphenating (716) said word between said pair of morphemes when said second type hyphenation location is in said preferred hyphenation area and said third type hyphenation location is not in said preferred hyphenation area;

hyphenating (716) said word between said pair of syllables when said first type hyphenation location is in said preferred hyphenation area, said second type hyphenation location is not in said preferred hyphenation area and said third type hyphenation location is not in said preferred hyphenation area.

5. A computer method for performing a ranked preferred hyphenation of words, comprising the steps of:

storing (802) a word in a computer memory, having at least a pair of adjacent syllables, a pair of adjacent morphemes and a hard hyphen;

storing (804) a first location value in said memory for a first type hyphenation location between said pair of syllables;

storing (806) a second location value in said memory for a second type hyphenation location between said pair of morphemes;

WEST

storing (810) a right margin value in said memory for defining a right margin of a document to be generated with a text processing program in said computer;

storing (812) a preferred zone value in said memory for defining a preferred hyphenation area left of said right margin of said document;

hyphenating (814) said word at said hard hyphen when said hard hyphen is detected in said preferred hyphenation area;

hyphenating (818) said word between said pair of morphemes when said second type hyphenation location is in said preferred hyphenation area and said hard hyphen is not detected in said preferred hyphenation area;

hyphenating (820) said word between said pair of syllables when said first type hyphenation location is in said preferred hyphenation area, said second type hyphenation location is not in said preferred hyphenation area and said hard hyphen is not detected in said preferred hyphenation area.

6. A computer method for performing a ranked preferred hyphenation of words, comprising the steps of:

storing (802) a word in a computer memory, having at least a pair of adjacent syllables, a pair of adjacent component words of a compound word and a hard hyphen;

storing (804) a first location value in said memory for a first type hyphenation location between said pair of syllables;

storing (806) a second location value in said memory for a second type hyphenation location between said pair of component words of a compound word;

storing (810) a right margin value in said memory for defining a right margin of a document to be generated with a text processing program in said computer;

storing (812) a preferred zone value in said memory for defining a preferred hyphenation area left of said right margin of said document;

hyphenating (814) said word at said hard hyphen when said hard hyphen is detected in said preferred hyphenation area;

hyphenating (816) said word between said pair of component words of a compound word when said second type hyphenation location is in said preferred hyphenation area and said hard hyphen is not detected in said preferred hyphenation area;

hyphenating (820) said word between said pair of syllables when said first type hyphenation location is in said preferred hyphenation area, said second type hyphenation location is not in said preferred hyphenation area and said hard hyphen is not detected in said preferred hyphenation area.

7. A computer method for performing a ranked preferred hyphenation of words, comprising the steps of:

storing (802) a word in a computer memory, having at least a pair of adjacent syllables, a pair of adjacent morphemes, a pair of adjacent component words of a compound word and a hard hyphen;

storing (804) a first location value in said memory for a first type hyphenation location between said pair of syllables;

storing (806) a second location value in said memory for a second type hyphenation location between said pair of morphemes;

storing (808) a third location value in said memory for a third type hyphenation location between said pair of component words of a compound word;

storing (810) a right margin value in said memory for defining a right margin of a document to be generated with a text processing program in said computer;

storing (812) a preferred zone value in said memory for defining a preferred hyphenation area left of said right margin of said document;

hyphenating (814) said word at said hard hyphen when said hard hyphen is detected in said preferred hyphenation area;

hyphenating (816) said word between said pair of component words of a compound word when said third type hyphenation location is in said preferred hyphenation area and said hard hyphen is not detected in said preferred hyphenation area;

hyphenating (818) said word between said pair of morphemes when said second type hyphenation location is in said preferred hyphenation area, said third type hyphenation location is not in said preferred hyphenation area and said hard hyphen is not detected in said preferred hyphenation area;

hyphenating (820) said word between said pair of syllables when said first type hyphenation location is in said preferred hyphenation area, said second type hyphenation location is not in said preferred hyphenation area, said third type hyphenation location is not in said preferred hyphenation area and said hard hyphen is not detected in said preferred hyphenation area.

8. A computer method for performing a ranked preferred hyphenation of words, comprising the steps of:

storing a word in a computer memory, having at least a pair of adjacent syllables and a pair of adjacent special processing segments;

storing a first location value in said memory for a first type hyphenation location between said pair of syllables;

storing a second location value in said memory for a second type hyphenation location between said pair of special processing segments;

storing a right margin value in said memory for defining a right margin of a document to be generated with a text processing program in said computer;

storing a preferred zone value in said memory for defining a preferred hyphenation area left of said right margin of said document;

hyphenating said word between said pair of special processing segments and performing special processing to at least one segment of said pair of segments when said second type hyphenation location is in said preferred hyphenation area;

hyphenating said word between said pair of syllables when said first type hyphenation location is in said preferred hyphenation area and said second type hyphenation location is not in said preferred hyphenation area.

WEST

FIG. 1

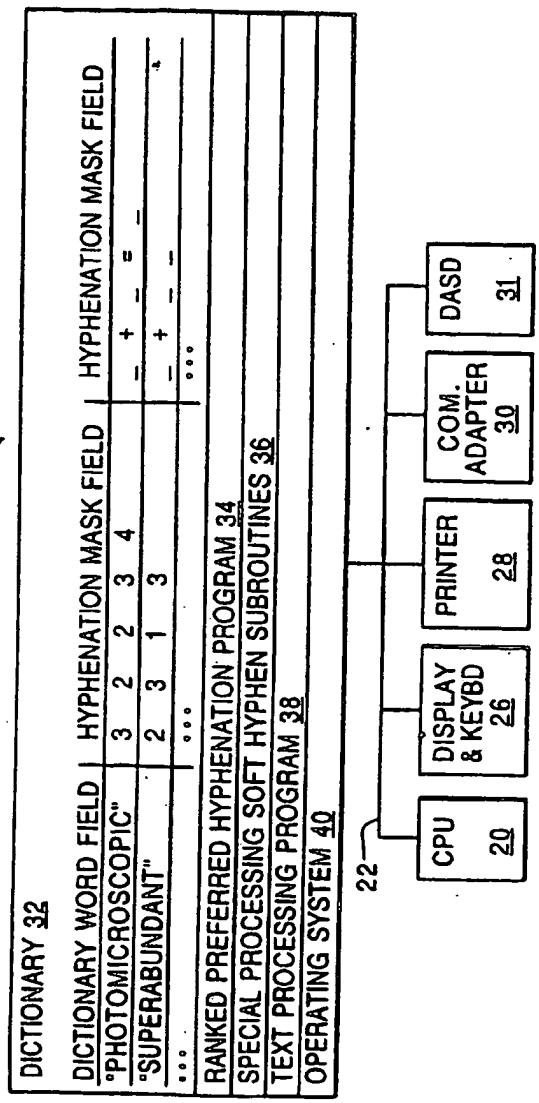

MEMORY 24
DICTIONARY 32
DICTIONARY WORD FIELD | HYPHENATION MASK FIELD | HYPHENATION MASK FIELD
"PHOTOMICROSCOPIC" 3 2 2 3 4 _ + _ = _
"SUPERABUNDANT" 2 3 1 3 _ + _ _
RANKED PREFERRED HYPHENATION PROGRAM 34
SPECIAL PROCESSING SOFT HYPHEN SUBROUTINES 36
TEXT PROCESSING PROGRAM 38
OPERATING SYSTEM 40
22
CPU 20
DISPLAY & KEYBD 26
PRINTER 28
COM. ADAPTER 30
DASD 31

FIG. 2

HYPHEN TYPE RANK (HIGH TO LOW)
- IS A HARD HYPHEN (ASCII CHARACTER 045)
+ IS A COMPONENT WORD BOUNDARY SOFT HYPHEN FOR A COMPOUND WORD
, IS A SPECIAL PROCESSING SOFT HYPHEN (NON-ENGLISH)
= IS A MORPHEME BOUNDARY SOFT HYPHEN
_ IS A SYLLABLE BOUNDARY SOFT HYPHEN
HIGH RANK
LOW RANK

FIG. 3

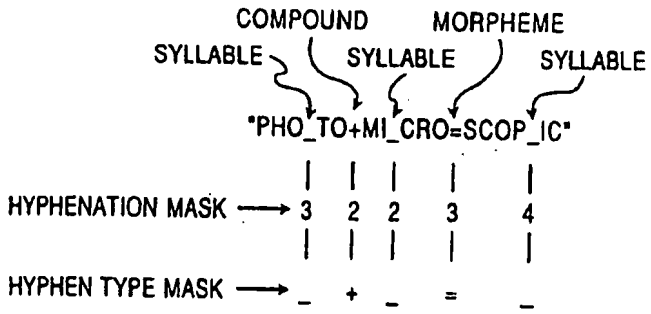
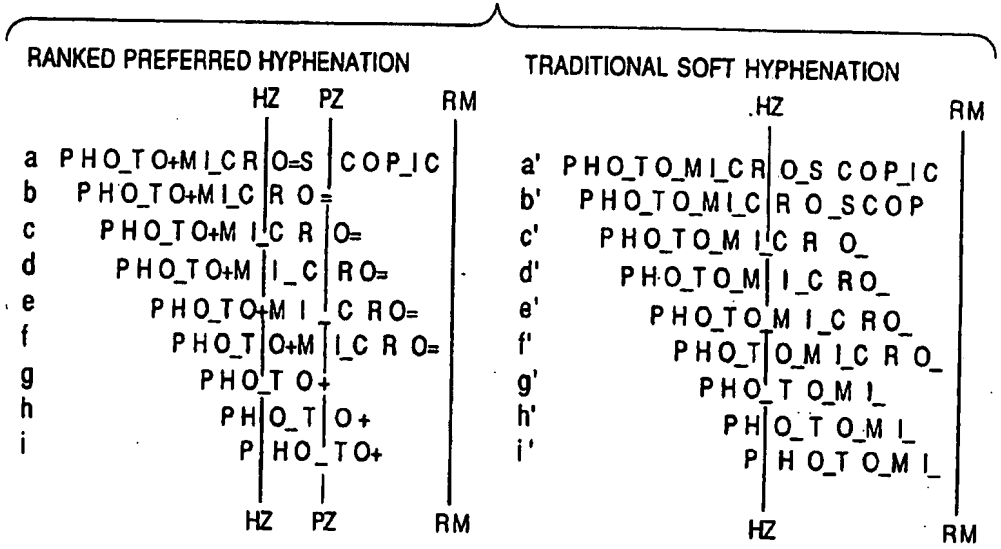


FIG. 4



WEST

FIG. 5

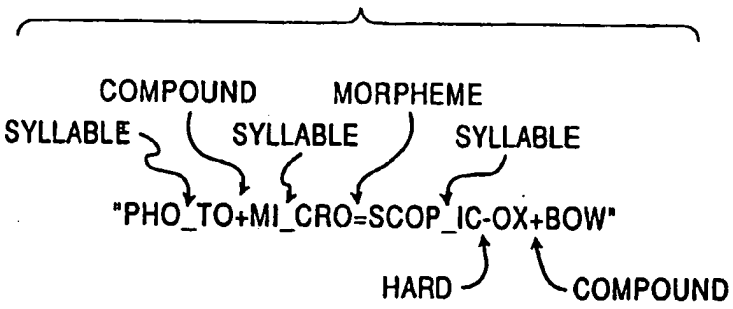

COMPOUND
MORPHEME
SYLLABLE
SYLLABLE
SYLLABLE
"PHO_TO+MI_CRO=SCOP_IC-OX+BOW"
HARD
COMPOUND

FIG. 6

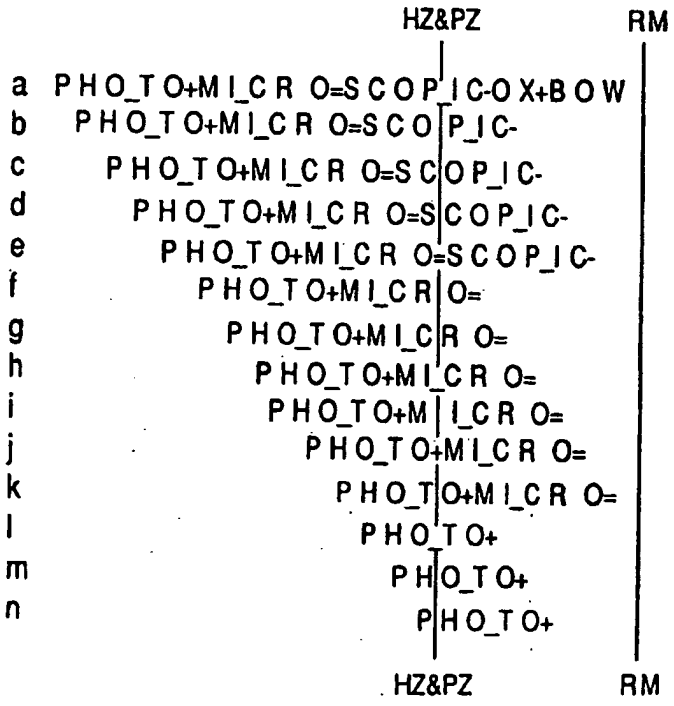
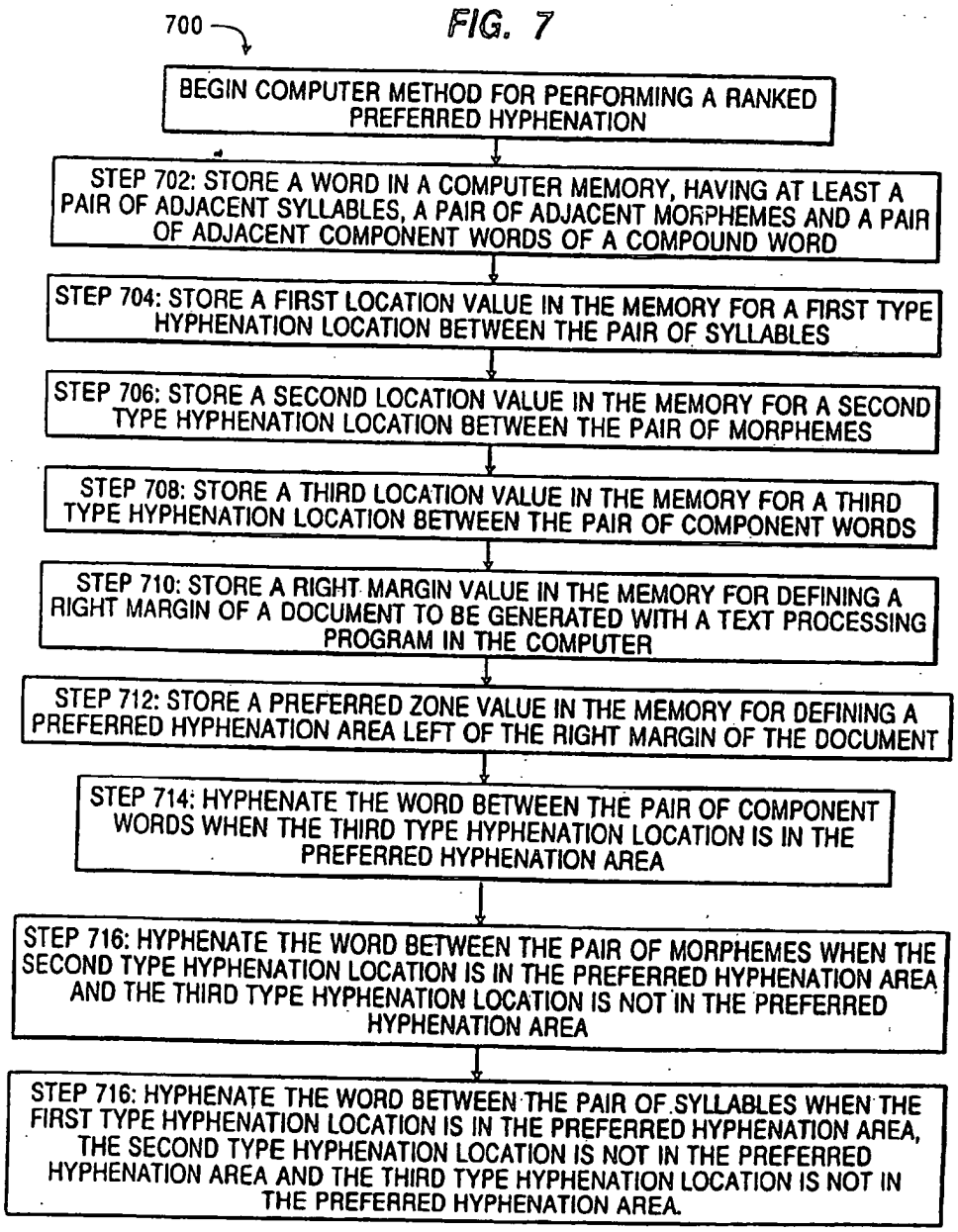
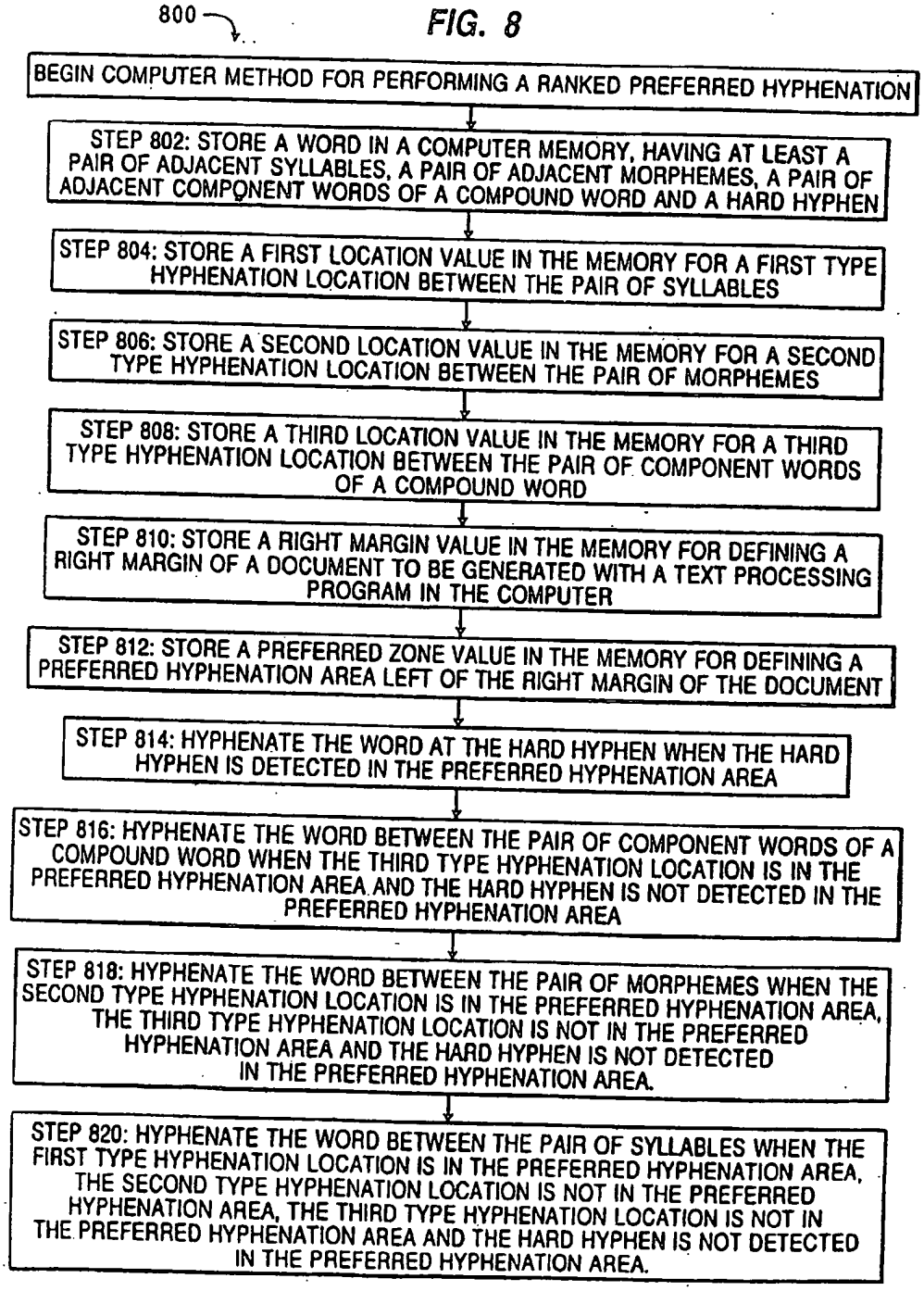

HZ&PZ
RM
a PHO_TO+MI_CR O=SCOP_IC-OX+BOW
b PHO_TO+MI_CR O=SCO P_IC-
c PHO_TO+MI_CR O=SC OP_IC-
d PHO_TO+MI_CR O=S COP_IC-
e PHO_TO+MI_CR O=SCOP_IC-
f PHO_TO+MI_CR O=
g PHO_TO+MI_CR O=
h PHO_TO+MI_CR O=
i PHO_TO+M I_CR O=
j PHO_TO+MI_CR O=
k PHO_TO+MI_CR O=
l PHO_TO+
m PHO_TO+
n PHO_TO+
HZ&PZ
RM

FIG. 7


700
BEGIN COMPUTER METHOD FOR PERFORMING A RANKED PREFERRED HYPHENATION
STEP 702: STORE A WORD IN A COMPUTER MEMORY, HAVING AT LEAST A PAIR OF ADJACENT SYLLABLES, A PAIR OF ADJACENT MORPHEMES AND A PAIR OF ADJACENT COMPONENT WORDS OF A COMPOUND WORD
STEP 704: STORE A FIRST LOCATION VALUE IN THE MEMORY FOR A FIRST TYPE HYPHENATION LOCATION BETWEEN THE PAIR OF SYLLABLES
STEP 706: STORE A SECOND LOCATION VALUE IN THE MEMORY FOR A SECOND TYPE HYPHENATION LOCATION BETWEEN THE PAIR OF MORPHEMES
STEP 708: STORE A THIRD LOCATION VALUE IN THE MEMORY FOR A THIRD TYPE HYPHENATION LOCATION BETWEEN THE PAIR OF COMPONENT WORDS
STEP 710: STORE A RIGHT MARGIN VALUE IN THE MEMORY FOR DEFINING A RIGHT MARGIN OF A DOCUMENT TO BE GENERATED WITH A TEXT PROCESSING PROGRAM IN THE COMPUTER
STEP 712: STORE A PREFERRED ZONE VALUE IN THE MEMORY FOR DEFINING A PREFERRED HYPHENATION AREA LEFT OF THE RIGHT MARGIN OF THE DOCUMENT
STEP 714: HYPHENATE THE WORD BETWEEN THE PAIR OF COMPONENT WORDS WHEN THE THIRD TYPE HYPHENATION LOCATION IS IN THE PREFERRED HYPHENATION AREA
STEP 716: HYPHENATE THE WORD BETWEEN THE PAIR OF MORPHEMES WHEN THE SECOND TYPE HYPHENATION LOCATION IS IN THE PREFERRED HYPHENATION AREA AND THE THIRD TYPE HYPHENATION LOCATION IS NOT IN THE PREFERRED HYPHENATION AREA
STEP 716: HYPHENATE THE WORD BETWEEN THE PAIR OF SYLLABLES WHEN THE FIRST TYPE HYPHENATION LOCATION IS IN THE PREFERRED HYPHENATION AREA, THE SECOND TYPE HYPHENATION LOCATION IS NOT IN THE PREFERRED HYPHENATION AREA AND THE THIRD TYPE HYPHENATION LOCATION IS NOT IN THE PREFERRED HYPHENATION AREA.

WEST

FIG. 8


800
BEGIN COMPUTER METHOD FOR PERFORMING A RANKED PREFERRED HYPHENATION
STEP 802: STORE A WORD IN A COMPUTER MEMORY, HAVING AT LEAST A PAIR OF ADJACENT SYLLABLES, A PAIR OF ADJACENT MORPHEMES, A PAIR OF ADJACENT COMPONENT WORDS OF A COMPOUND WORD AND A HARD HYPHEN
STEP 804: STORE A FIRST LOCATION VALUE IN THE MEMORY FOR A FIRST TYPE HYPHENATION LOCATION BETWEEN THE PAIR OF SYLLABLES
STEP 806: STORE A SECOND LOCATION VALUE IN THE MEMORY FOR A SECOND TYPE HYPHENATION LOCATION BETWEEN THE PAIR OF MORPHEMES
STEP 808: STORE A THIRD LOCATION VALUE IN THE MEMORY FOR A THIRD TYPE HYPHENATION LOCATION BETWEEN THE PAIR OF COMPONENT WORDS OF A COMPOUND WORD
STEP 810: STORE A RIGHT MARGIN VALUE IN THE MEMORY FOR DEFINING A RIGHT MARGIN OF A DOCUMENT TO BE GENERATED WITH A TEXT PROCESSING PROGRAM IN THE COMPUTER
STEP 812: STORE A PREFERRED ZONE VALUE IN THE MEMORY FOR DEFINING A PREFERRED HYPHENATION AREA LEFT OF THE RIGHT MARGIN OF THE DOCUMENT
STEP 814: HYPHENATE THE WORD AT THE HARD HYPHEN WHEN THE HARD HYPHEN IS DETECTED IN THE PREFERRED HYPHENATION AREA
STEP 816: HYPHENATE THE WORD BETWEEN THE PAIR OF COMPONENT WORDS OF A COMPOUND WORD WHEN THE THIRD TYPE HYPHENATION LOCATION IS IN THE PREFERRED HYPHENATION AREA AND THE HARD HYPHEN IS NOT DETECTED IN THE PREFERRED HYPHENATION AREA
STEP 818: HYPHENATE THE WORD BETWEEN THE PAIR OF MORPHEMES WHEN THE SECOND TYPE HYPHENATION LOCATION IS IN THE PREFERRED HYPHENATION AREA, THE THIRD TYPE HYPHENATION LOCATION IS NOT IN THE PREFERRED HYPHENATION AREA AND THE HARD HYPHEN IS NOT DETECTED IN THE PREFERRED HYPHENATION AREA.
STEP 820: HYPHENATE THE WORD BETWEEN THE PAIR OF SYLLABLES WHEN THE FIRST TYPE HYPHENATION LOCATION IS IN THE PREFERRED HYPHENATION AREA, THE SECOND TYPE HYPHENATION LOCATION IS NOT IN THE PREFERRED HYPHENATION AREA, THE THIRD TYPE HYPHENATION LOCATION IS NOT IN THE PREFERRED HYPHENATION AREA AND THE HARD HYPHEN IS NOT DETECTED IN THE PREFERRED HYPHENATION AREA.

WEST

FIG. 9

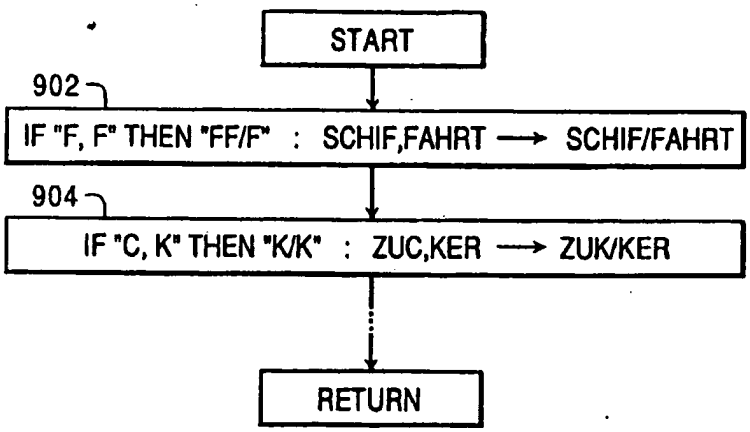
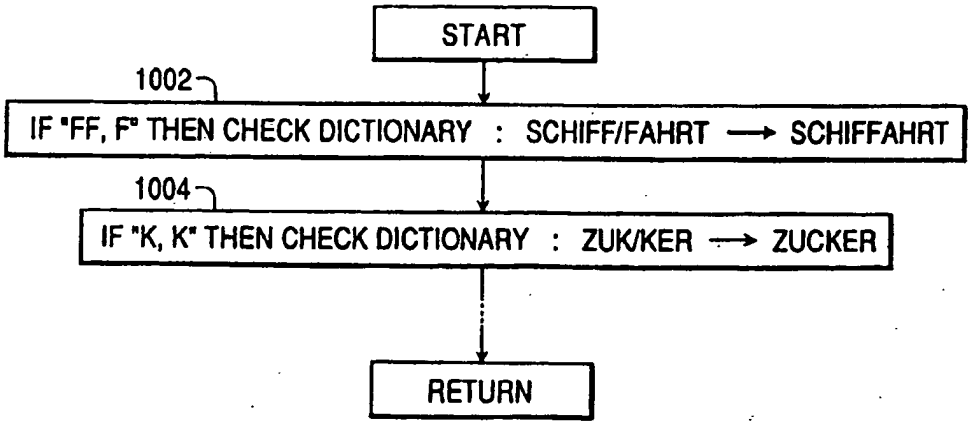
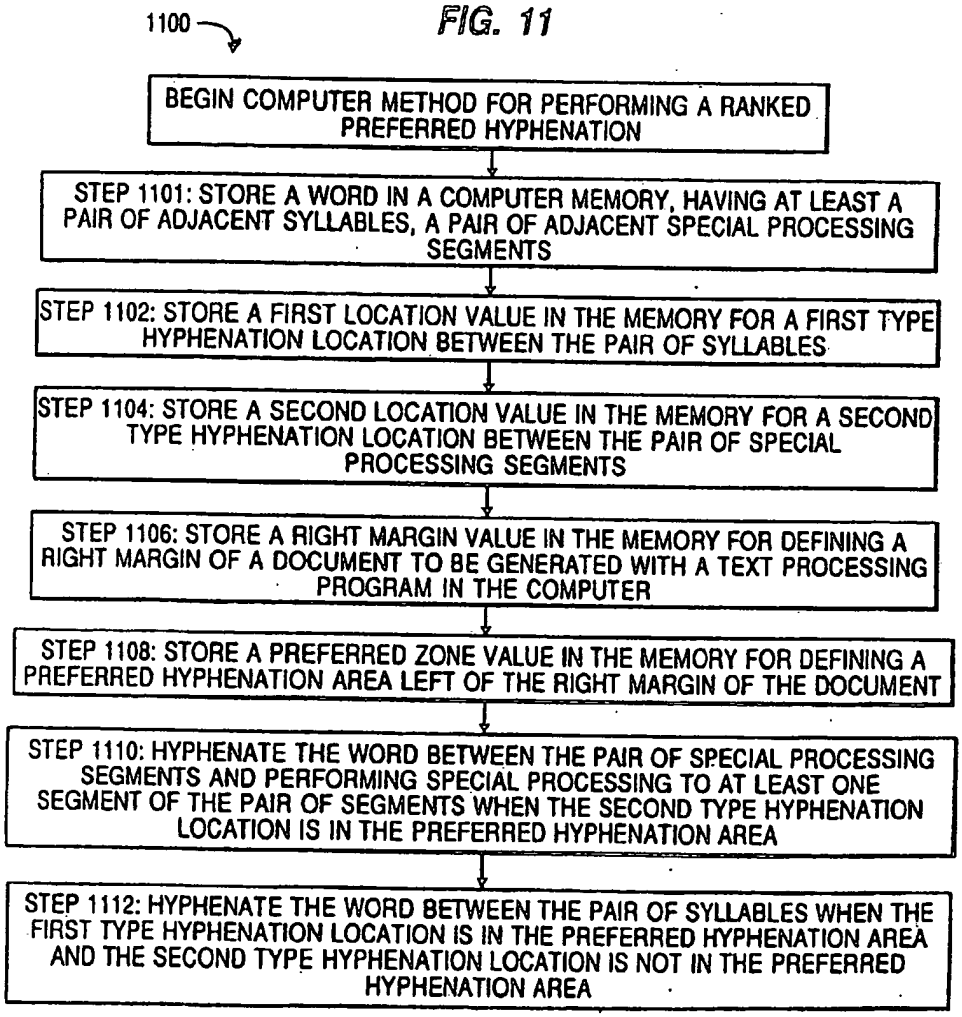


FIG. 10



WEST

FIG. 11


1100
BEGIN COMPUTER METHOD FOR PERFORMING A RANKED PREFERRED HYPHENATION
STEP 1101: STORE A WORD IN A COMPUTER MEMORY, HAVING AT LEAST A PAIR OF ADJACENT SYLLABLES, A PAIR OF ADJACENT SPECIAL PROCESSING SEGMENTS
STEP 1102: STORE A FIRST LOCATION VALUE IN THE MEMORY FOR A FIRST TYPE HYPHENATION LOCATION BETWEEN THE PAIR OF SYLLABLES
STEP 1104: STORE A SECOND LOCATION VALUE IN THE MEMORY FOR A SECOND TYPE HYPHENATION LOCATION BETWEEN THE PAIR OF SPECIAL PROCESSING SEGMENTS
STEP 1106: STORE A RIGHT MARGIN VALUE IN THE MEMORY FOR DEFINING A RIGHT MARGIN OF A DOCUMENT TO BE GENERATED WITH A TEXT PROCESSING PROGRAM IN THE COMPUTER
STEP 1108: STORE A PREFERRED ZONE VALUE IN THE MEMORY FOR DEFINING A PREFERRED HYPHENATION AREA LEFT OF THE RIGHT MARGIN OF THE DOCUMENT
STEP 1110: HYPHENATE THE WORD BETWEEN THE PAIR OF SPECIAL PROCESSING SEGMENTS AND PERFORMING SPECIAL PROCESSING TO AT LEAST ONE SEGMENT OF THE PAIR OF SEGMENTS WHEN THE SECOND TYPE HYPHENATION LOCATION IS IN THE PREFERRED HYPHENATION AREA
STEP 1112: HYPHENATE THE WORD BETWEEN THE PAIR OF SYLLABLES WHEN THE FIRST TYPE HYPHENATION LOCATION IS IN THE PREFERRED HYPHENATION AREA AND THE SECOND TYPE HYPHENATION LOCATION IS NOT IN THE PREFERRED HYPHENATION AREA

WEST